Cell, Vol. 80, 825–827, March 24, 1995, Copyright © 1995 by Cell Press

"If the Shoe Fits": Clues on Structural Recognition of DNA Damage

Minireview

James E. Cleaver and Susan K. Layher
Laboratory of Radiobiology and Environmental Health
University of California, San Francisco
San Francisco, California 94143-0750

The study of DNA repair mechanisms in prokaryotic and eukaryotic cells has matured greatly in the last few years. Genes have been cloned, proteins have been isolated, and mechanisms are being defined (Sancar, 1994; Hanawalt, 1994). It is now possible to address fundamental questions that lie at the heart of DNA repair. How do cells recognize specific lesions in the context of the large excess of undamaged DNA and of the various metabolic processes that can occur simultaneously in the same regions of DNA? The ability of proteins to locate a specific sequence or structural aberration within a large excess of undamaged DNA is crucial in regulating such cellular functions as gene expression, initiation of DNA synthesis, genetic recombination, restriction/modification, and DNA repair. The studies reported by Mol et al. (1995 [this issue of *Cell*]) and Savva et al. (1995) on the crystal structure and mechanism of uracil-DNA glycosylase (UDG) from human and viral sources represent important advances in understanding how proteins recognize atypical structures in DNA. UDG carries out all the necessary functions of repair—scanning DNA, locating damaged bases with high discrimination, specific binding, and catalysis—within a small, single polypeptide. The mechanism of recognition may consequently be a paradigm for many other recognition mechanisms in which high specificity is required for DNA repair.

Repair systems in general exhibit a wide range of abilities, from exquisite specificity to great versatility. Base excision repair, typified by UDG, involves a family of highly specific monofunctional glycosylases, each tailored to the excision of defined modified bases such as uracil, hypoxanthine, formamidopyrimidine, urea, thymine glycol, 3-methyladenine, and 7-methylguanine (Sancar and Sancar, 1988). The glycosylases' recognition mechanisms detect and discriminate between normal and inappropriate or defective bases that produce minimal disruption in base pairing and stacking. Recognition of the damaged base, binding to DNA, and cleavage of the glycosyl bond are performed by small monomeric proteins, most of which require no divalent cations, prefer double-stranded DNA, and typically are of 20–30 kDa in size (Sancar and Sancar, 1988). Nucleotide excision repair, in contrast, is performed by protein complexes composed of separate recognition proteins, helicases, and nucleases that excise nucleotide segments of DNA around almost any damaged site. A bewildering variety of simple lesions, photoproducts, and bulky chemical adducts can be removed by this mechanism that appears to pay little attention to the specific kind of DNA modification involved (Sancar, 1994; Grossman, 1994).

Recent structural analysis of UDG (Mol et al., 1995; Savva et al., 1995) and of other enzymes with glycosylase activity (Kuo et al., 1992; Morikawa et al., 1992, 1994; Latham and Lloyd, 1994) suggests that there may be at least three series of glycosylases, distinguished by their potential mechanisms of action and substrate specificity. The first series, represented by UDG, acts on naturally occurring deamination products or misincorporated bases such as uracil, thymine, and hypoxanthine. A second series acts on damaged bases such as 8-oxoguanine, 7-methylguanine, 3-methyladenine, urea, and other derivatives from ring-opened pyrimidines and other oxidative products (Sancar and Sancar, 1988). The third series includes bifunctional enzymes that have a glycosylase and an endonuclease activity in one protein. Typical of these are the pyrimidine dimer glycosylases from the T4 phage (T4denV) and Micrococcus luteus (Morikawa et al., 1992, 1994; Latham and Lloyd, 1994), as well as endonuclease III (endo III), which excises oxidized bases (Kuo et al., 1992).

Uracil in DNA

Cytosine deamination, one of the more frequent chemical reactions known to create serious DNA damage in cells, can occur at a significant rate by spontaneous hydrolysis under physiological conditions. At neutral pH, cytosine can be deaminated to uracil by either hydrolytic deamination via protonation of the amino group or direct hydroxyl ion attack on the C-4 of the pyrimidine ring. Uracil can therefore be produced in DNA directly or can be misincorporated as a deaminated product from the DNA precursor pool. Uracil in DNA presents a unique problem for DNA repair processes since it is found normally in RNA and in certain circumstances, even in DNA. If unrepaired, it can give rise to GC to AT transitions. Escherichia coli strains lacking glycosylase activity (*ung*) are viable but show elevated spontaneous and induced mutation rates and sensitivity to chemicals such as bisulfite and nitrous acid, which cause cytosine deamination. The existence of UDG and the discovery of base excision repair itself came from a direct search in cell extracts for an enzyme that corrects deaminated cytosine residues in DNA (Lindahl, 1974). Subsequently, it was found that uracil-containing Bacillus subtilis phages, PBS1 and PBS2, produce an inhibitor of the host's UDG, allowing for phage replication (Friedberg et al., 1975). UDG activity has since been found in extracts of prokaryotes, lower eukaryotes, and a variety of mammalian cells and tissues, as well as in many viruses (Savva et al., 1995; Duncan, 1981). The enzyme is essential for maintaining genomic stability in all DNA-containing organisms and may play a particular role during viral reactivation, repairing cytosine deaminations accumulated during viral latency.

UDG—A Swing Out Mechanism

UDG, a 20–30 kDa protein, acts on both single-stranded and double-stranded DNA. Excision from single-stranded DNA is more rapid and less dependent on flanking bases

than is excision from double-stranded DNA. Recent crystal structure analysis of the human enzyme by Mol et al. (1995) and of the virally encoded enzyme from herpes simplex 1 (HSV-1) by Savva et al. (1995) represents an important advance in understanding molecular mechanisms of lesion recognition. The human enzyme was cocrystallized with an inhibitor, 6-aminouracil, bound in its active site, and either the HSV-1 enzyme was cocrystallized with a thymine trinucleotide or the crystals were soaked with free uracil to identify both DNA-binding and active sites. These two approaches in large part support a mechanism by which the protein binds single-stranded or double-stranded DNA in a groove and causes the uracil to be swung out into an extrahelical position to fit into the active site. Together, details presented by both approaches convey a novel picture of binding and catalysis for base excision.

The general structures of both human and viral molecules are closely similar. The viral enzyme is 39% identical to the human enzyme, with the C-terminus being more conserved. The main variability is seen in the N-terminus, which may be involved in other functions, such as cellular localization. Both enzymes are compact, flat, and elliptical with an α/β fold that shapes the enzyme's DNA-binding site in the form of a positively charged, tapered groove surrounded by a high local concentration of rigid proline residues. This groove has a width of about 20 Å that corresponds to the DNA helix at one end, but narrows to about 10 Å at the other. Savva et al. (1995) noted that a closer fit can be obtained by slight bending of double-stranded DNA onto the positively charged face of the protein. In addition, their analysis of thymine trinucleotide cocrystallized with UDG suggests a more disordered conformation for single-stranded DNA that nevertheless also binds closely to the groove. The overall structure of UDG is aptly described as "resembling a slightly dented matchbox" (Savva et al., 1995). The groove may permit both DNA binding and a limited amount of scanning until a uracil is located, but the absence of cofactors or ATPase activity limits the extent of scanning possible as compared with the $UvrA_2B$ excision complex, for example (Grossman, 1994).

Of interest is the report of an analogous enzyme involved in correcting GT mismatches, thymine-DNA glycosylase, which is more than twice as large as UDG and has an absolute requirement for double-stranded DNA (Jiricny, 1994; Neddermann and Jiricny, 1994). The relatively larger size of thymine-DNA glycosylase might be relevant to its need to recognize the opposing strand since an incorrectly positioned thymine in single-stranded DNA would be unrecognizable.

The most striking feature of the UDG structure discovered by both Mol et al. (1995) and Savva et al. (1995) is an active site that requires the uracil to become extrahelical. At the base of the groove, there is an internal recognition pocket for uracil that excludes other pyrimidine and purine bases. The active site is also identified on the basis of inactivating mutations and sequence conservation of several residues (especially critical aspartic acid, asparagine, and histidine residues) at one end of the groove (Mol et al., 1995). Binding of 6-aminouracil or uracil in the two crystal structures indicates how an extrahelical uracil base can be created, recognized, and removed.

For the long axis of duplex DNA parallel to the groove, multiple hydrogen bonds can form between the enzyme's recognition pocket and the extrahelical uracil ring at positions 2, 3, and 4. The study using thymine trinucleotide shows that an additional "trap" may hold the thymine base out of the pocket and provide additional discrimination (Savva et al., 1995). This recognition mechanism involving an extrahelical base has been observed previously only for the HhaI (cytosine-5) methyltransferase that was crystallized with a 13 bp oligonucleotide substrate (Klimasauskas et al., 1994). In that case, the displacement of cytosine appeared to be facilitated by a peptide loop of the protein, which is not evident in the UDG. The UDG's recognition pocket may instead serve to capture uracil in an extrahelical configuration, which may occur spontaneously.

The rotation of the uracil brings the C-1′ atom of the deoxyribose residue within range of nucleophilic attack. In RNA, the 2′-OH should interfere sterically with attack on the C-1′ atom. Although the specificity of the human and viral UDGs is identical, there are subtle differences in the proposed mechanisms of the two nucleophilic attacks involving the absolutely conserved histidine and aspartate residues in the active site. In the human enzyme, the histidine residue interacts with a water molecule to abstract a proton and attack the C-1′ position. Yet, in the HSV-1 enzyme, it is the aspartate residue that abstracts a proton from a water molecule, which then attacks the C-1′ atom. In the human enzyme, main chain amides of the aspartate residue and its neighboring glutamine residue then form an oxyanion at the O-2 position of uracil. However, in the HSV-1 enzyme, it is the histidine residue that attacks the O-2 position of uracil. The end result for both proposed mechanisms is the release of uracil, which may remain temporarily bound to the enzyme's pocket and inhibit the enzyme.

Damage-Specific Glycosylases

A second series of glycosylases are known with specificities for bases damaged by alkylation or oxidative reactions (Sancar and Sancar, 1988). The enzyme substrates may be continuously formed by nonenzymatic alkylation of DNA from S-adenosylmethionine under physiological conditions and by a range of oxidative processes that generate 8-oxoguanine as a major product in both nuclear and mitochondrial DNA. Alkylations at DNA base positions, N3 adenine and N7 guanine, differ from the deamination products, uracil and thymine, in the sense that the alkyl groups added to these purines protrude into the minor and major grooves of the DNA, respectively. Unless repaired, these bases may become reactive electrophiles by virtue of exposed positive charge, their structural motifs, or both. Whereas UDG has a strict substrate specificity, this series of glycosylases has slightly broader ranges on damaged substrates. Although it is tempting to generalize the structural information obtained from UDG, the substrate ranges

of the damage-specific glycosylases and the presence of numerous structural and charge changes on DNA bases suggest that recognition mechanisms of these glycosylases could differ significantly. The rigid pocket seen in UDG would not be appropriate for enzymes with a broader substrate range.

Two other glycosylases have been studied at the crystal structure level, endo III (Kuo et al., 1992) and the T4denV (Morikawa et al., 1992, 1994). These enzymes differ from UDG in having two enzymatic activities that reside in a single protein. E. coli endo III is a 23 kDa DNA iron–sulfur glycosylase that removes ring-saturated, ring-cleaved, and ring-contracted pyrimidines, such as thymine glycols, from oxidized and X-irradiated DNA. The enzyme also possesses a divalent cation-independent AP endonuclease activity that incises the phosphodiester backbone 3′ to the AP site by β elimination (Kuo et al., 1992). T4 endoV is a 16 kDa pyrimidine dimer-specific DNA glycosylase responsible for the hydrolysis of the 5′-glycosyl bond of a *cis-syn* pyrimidine dimer and the subsequent abasic site to generate an α/β unsaturated aldehyde and a 5′-phosphomonoester terminus by a β-elimination mechanism (Morikawa et al., 1994; Latham and Lloyd, 1994). Structural analysis suggests that the T4denV binds in the minor groove of B-DNA, employing a region of the protein that is mobile in solution but may form a defined tertiary structure upon binding (Morikawa et al., 1992, 1994). No evidence for an extrahelical dimer analogous to the mechanism of UDG has been found. The DNA binding and the bifunctionality of both endo III and T4denV appear to be achieved by a flexible fit of the proteins to the damaged residues in DNA, unlike UDG, for which the DNA base undergoes the major conformational change.

Other Damage-Specific Recognition Mechanisms

On the basis of these studies, what other motifs are likely to be discovered for proteins that recognize different kinds of DNA damage? Repair of cyclobutane pyrimidine dimers has been a dominant theme in the development of our concepts of DNA repair. Currently, there may be up to six distinct kinds of proteins that can carry out dimer recognition, with wide degrees of specificity. These include class I (microbial) and class II (higher eukaryotic) photolyases (Yasui et al., 1994), T4denV (Morikawa et al., 1992, 1994), $UvrA_2B$ of E. coli (Grossman, 1994), and two mammalian recognition proteins, XPA and XPE (Jones and Wood, 1993; Sancar, 1994; Reardon et al., 1993). Structural information is currently only available for T4denV (Morikawa et al., 1992, 1994), although there has been considerable progress made on the crystallization and analysis of class I photolyases (Park et al., 1993). The proteins of higher order repair systems show more versatility than the highly specific glycosylases; however, the resolution of their structure and mechanisms of action may have some interesting parallels to the glycosylases.

The studies of Mol et al. (1995) and Savva et al. (1995) on UDG have consequently revealed unexpected recognition mechanisms with an extrahelical base conformation that will undoubtedly be influential in stimulating heuristic thought and experiments about fundamental recognition mechanisms that are central to DNA repair for both eukaryotic and prokaryotic species.

References

Duncan, B. K. (1981). In The Enzymes: Nucleic Acids, Part A, P. D. Boyer, ed. (New York: Academic Press), 565–586.

Friedberg, E. C., Ganesan, A. K., and Minton, K. (1975). J. Virol. *16*, 315–321.

Grossman, L. (1994). Ann. NY Acad. Sci. *726*, 252–266.

Hanawalt, P. C. (1994). Science *266*, 1957–1958.

Jiricny, J. (1994). Trends Genet. *10*, 164–168.

Jones, C. J., and Wood, R. D. (1993). Biochemistry *32*, 12096–12104.

Klimasauskas, S., Kumar, S., Roberts, R. J., and Cheng, X. (1994). Cell *76*, 357–369.

Kuo, C.-F., McRee, D. E., Fisher, C. L., O'Handley, S. F., Cunningham, R. P., and Tainer, J. A. (1992). Science *258*, 434–440.

Latham, K. A., and Lloyd, R. S. (1994). Ann. NY Acad. Sci. *726*, 181–197.

Lindahl, T. (1974). Proc. Natl. Acad. Sci. USA *71*, 3649–3653.

Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E., and Tainer, J. A. (1995). Cell *80*, this issue.

Morikawa, K., Matsumoto, O., Tsujimoto, M., Katayanagi, K., Ariyoshi, M., Doi, T., Ikehira, M., Inaoka, T., and Ohtsuka, E. (1992). Science *256*, 523–526.

Morikawa, K., Ariyoshi, M., Vassylyev, D., Katayanagi, K., Nakamura, H., Doi, T., Hori, N., and Ohtsuka, E. (1994). Ann. NY Acad. Sci. *726*, 198–207.

Neddermann, P., and Jiricny, J. (1994). Proc. Nat. Acad. Sci. USA *91*, 1642–1646.

Park, H.-W., Sancar, A., and Deisenhofer, J. (1993). J. Mol. Biol. *231*, 1122–1125.

Reardon, J. T., Nichols, A. F., Keeney, S., Smith, C. A., Taylor, J. S., Linn, S., and Sancar, A. (1993). J. Biol. Chem. *268*, 21301–21308.

Sancar, A. (1994). Science *266*, 1954–1956.

Sancar, A., and Sancar, G. B. (1988). Annu. Rev. Biochem. *57*, 29–67.

Savva, R., MacAuley-Hecht, K., Brown, T., and Pearl, L. (1995). Nature *373*, 487–493.

Yasui, A., Eker, A. P. M., Yasuhira, S., Yajima, H., Kobayashi, T., Takao, M., and Oikawa, A. (1994). EMBO J. *13*, 6143–6151.